

MetaMath: Bootstrap Your Own Mathematical Questions for Large Language Models

Year: 2023 | Citations: 533 | Authors: L. Yu, Weisen Jiang, Han Shi

Abstract

Large language models (LLMs) have pushed the limits of natural language understanding and exhibited excellent problem-solving ability. Despite the great success, most existing open-source LLMs (e.g., LLaMA-2) are still far away from satisfactory for solving mathematical problem due to the complex reasoning procedures. To bridge this gap, we propose MetaMath, a fine-tuned language model that specializes in mathematical reasoning. Specifically, we start by bootstrapping mathematical questions by rewriting the question from multiple perspectives without extra knowledge, which results in a new dataset called MetaMathQA. Then we fine-tune the LLaMA-2 models on MetaMathQA. Experimental results on two popular benchmarks (i.e., GSM8K and MATH) for mathematical reasoning demonstrate that MetaMath outperforms a suite of open-source LLMs by a significant margin. Our MetaMath-7B model achieves 66.4% on GSM8K and 19.4% on MATH, exceeding the state-of-the-art models of the same size by 11.5% and 8.7%. Particularly, MetaMath-70B achieves an accuracy of 82.3% on GSM8K, slightly better than GPT-3.5-Turbo. We release all the MetaMathQA dataset, the MetaMath models with different model sizes and the training code for public use.